

***N*-Jettiness: An Inclusive Event Shape to Veto Jets**

Iain W. Stewart, Frank J. Tackmann, and Wouter J. Waalewijn

Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139

Jet vetoes are essential in many Higgs and new-physics analyses at the LHC and Tevatron. The signals are typically characterized by a specific number of hard jets, leptons, or photons, while the backgrounds often have additional jets. In such cases vetoing undesired additional jets is an effective way to discriminate signals and background. Given an inclusive event sample with N or more jets, the veto to have only N energetic jets defines an “exclusive” N -jet cross section. This strongly restricts the phase space of the underlying inclusive N -jet cross section and causes large double logarithms in perturbation theory that must be summed to obtain theory predictions. Jet vetoes are typically implemented using jet algorithms. This yields complicated phase-space restrictions and one often relies on parton-shower Monte Carlos, which are limited to leading-logarithmic accuracy. We introduce a global event shape “ N -jettiness”, τ_N , which is defined for events with N signal jets and vanishes in the limit of exactly N infinitely narrow jets. Requiring $\tau_N \ll 1$ constrains radiation between the N signal jets and vetoes additional undesired jets. This provides an inclusive method to veto jets and to define an exclusive N -jet cross section that can be well-controlled theoretically. N -jettiness yields a factorization formula with inclusive jet and beam functions.

Introduction. At the LHC or Tevatron, hard interactions involving Higgs or new-physics particles are identified by looking for signals with a characteristic number of energetic jets, leptons, or photons [1]. The backgrounds come from Standard Model processes producing the same signature of hard objects possibly with additional jets. An example are top quarks decaying into W plus b -jet, which is a major background for $H \rightarrow WW$ [2]. When reconstructing masses and decay chains of new-physics particles additional jets can cause large combinatorial backgrounds. Standard Model processes can also fake a signal when a jet is misidentified as lepton or photon, a typical example being $H \rightarrow \gamma\gamma$.

Thus, a veto on additional undesired jets is an effective and sometimes necessary method to clean up the events and discriminate signal and the various backgrounds. More generally, one would like to measure an “exclusive” N -jet cross section, $pp \rightarrow XL(Nj)$, to produce N signal jets j where the remaining X contains no hard (central) jets. Here, $N \geq 0$ and L denotes the hard leptons or photons required as part of the signal.

We introduce an inclusive event shape “ N -jettiness”, denoted τ_N and defined below in Eq. (1). For an event with at least N energetic jets, τ_N provides an inclusive measure of how N -jet-like the event looks. In the limit $\tau_N \rightarrow 0$ the event contains exactly N infinitely narrow jets. For $\tau_N \sim 1$ the event has hard radiation between the N signal jets. Requiring $\tau_N \ll 1$ constrains the radiation outside the signal and beam jets, providing an inclusive way to veto additional central jets. It yields an inclusive definition of an exclusive N -jet cross section with a smooth transition between the case of no jet veto, $\tau_N \sim 1$, and the extremely exclusive case $\tau_N \rightarrow 0$.

Vetoing additional jets imposes a phase-space restriction on the underlying inclusive N -jet cross section to produce N or more jets with the same L . Irrespective of its precise definition, the jet veto introduces a jet resolu-

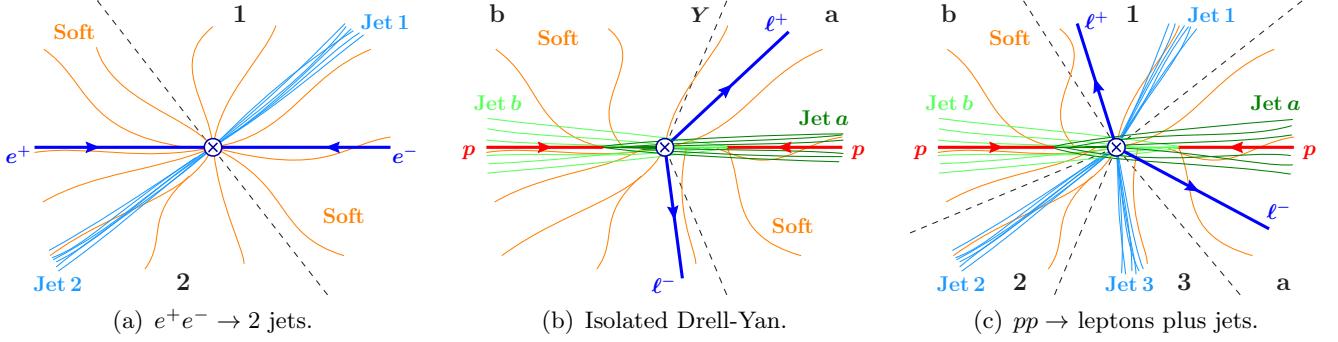
tion scale μ_J that characterizes this restriction, i.e. the distinction between N and $N+1$ jets. Hence, the exclusive N -jet cross section contains phase-space logarithms $\alpha_s^n \ln^m(\mu_J^2/\mu_H^2)$, where $m \leq 2n$ and μ_H is the scale of the hard interaction. For τ_N , $\mu_J^2/\mu_H^2 \simeq \tau_N \ll 1$. Generically there is always a hierarchy $\mu_J \ll \mu_H$, which becomes larger the stronger the restrictions are. These large logarithms must be summed to obtain reliable predictions.

Jet vetoes are typically implemented by using a jet algorithm to find all jets in the event and veto events with too many energetic jets. Jet algorithms are good tools to identify the signal jets. However, they are not necessarily well-suited to veto unwanted jets, because the corresponding phase-space restrictions are complicated and depend in detail on the algorithm. This makes it difficult to incorporate the jet veto into explicit theoretical calculations and inhibits a systematic summation of the resulting large logarithms. In this case, usually the only way to predict the corresponding exclusive N -jet cross section is to rely on parton shower Monte Carlos to sum the leading logarithms (LL). For particular jet algorithms, the resolution y_{23} defines the transition from 2 to 3 jets. Next-to-leading logarithms for this and other hadron-collider event shapes were summed in Ref. [3].

Vetoing jets by cutting on an inclusive variable like τ_N has several advantages. First, we can go beyond LL order, because the logarithms from the phase-space restriction, $\alpha_s^n \ln^m \tau_N$, are simple enough to allow their systematic summation to higher orders. Moreover, the theory predictions with factorization can be directly compared with experiment without having to utilize Monte Carlos for parton showering or hadronization. Experimentally, τ_N reduces the dependence on jet algorithms and might help improve the background rejection.

Definition. N -jettiness is defined as

$$\tau_N = \frac{2}{Q^2} \sum_k \min\{q_a \cdot p_k, q_b \cdot p_k, q_1 \cdot p_k, \dots, q_N \cdot p_k\}. \quad (1)$$

FIG. 1: Different situations for the application of N -jettiness.

As we discuss below, this definition of τ_N yields a factorization formula with inclusive jet and beam functions and allows the summation of logarithms to next-to-next-to-leading logarithmic (NNLL) order. The sum over k in Eq. (1) runs over the momenta p_k of all measured (pseudo-)particles in the final state excluding the signal leptons or photons in L . (Any other leptons or photons, e.g. from hadronic decays, are included in the sum.) For simplicity we take all p_k to be massless. The q_a , q_b , and q_1, \dots, q_N are a fixed set of massless reference momenta for the two beams and the N signal jets,

$$q_{a,b}^\mu = \frac{1}{2} x_{a,b} E_{\text{cm}} n_{a,b}^\mu, \quad n_a^\mu = (1, \hat{z}), \quad n_b^\mu = (1, -\hat{z}),$$

$$q_J^\mu = E_J(1, \hat{n}_J), \quad J = \{1, \dots, N\}. \quad (2)$$

The E_J and \hat{n}_J correspond to the energies and directions of the N signal jets (for both massive and massless jets). Their choice is discussed below. The beam reference momenta q_a and q_b are the large momentum components of the colliding partons along the beam axis (taken to be the z axis). They are defined by

$$x_a E_{\text{cm}} = n_b \cdot (q_1 + \dots + q_N + q), \quad (3)$$

and analogously for x_b with $a \leftrightarrow b$. Here, q is the total momentum of the non-hadronic signal L . In Eq. (1), $Q^2 = x_a x_b E_{\text{cm}}^2$ is the hard interaction scale, and the distance of a particle with momentum p_k from the jets or beams is measured by $q_m \cdot p_k$. If L contains missing energy, so q and $x_{a,b}$ are not known, one can use a modified distance measure as we discuss below Eq. (11).

The minimum for each k in Eq. (1) associates the particle with the closest beam or jet, appropriately dividing the hadronic initial-state radiation (ISR) and final-state radiation (FSR). Soft particles and energetic particles near any jet or beam only give small contributions to the sum. For $2 \rightarrow N$ scattering of massless partons, $\tau_N = 0$. Energetic particles far away from all jets and beams give large contributions. Hence, for $\tau_N \ll 1$ the final state has N jets, two forward beam jets, and only soft radiation between them. In this limit $x_{a,b}$ are the momentum fractions of the annihilated partons, and $Y = \ln(x_a/x_b)/2$ is the boost of the partonic center-of-mass frame.

$N = 2$ for $e^+e^- \rightarrow \text{jets}$. In e^+e^- collisions there is no hadronic ISR, so we drop the $q_{a,b} \cdot p_k$ entries in Eq. (1). Now Q^2 is the total invariant mass of the leptons and $Y = 0$. In the two-jet limit, the jet directions are close to the thrust axis \hat{t} , defined by the thrust $T = \max_{\hat{t}} \sum_i |\hat{t} \cdot \vec{p}_i|/Q$. Hence we can choose

$$q_1^\mu = \frac{1}{2} Q(1, \hat{t}), \quad q_2^\mu = \frac{1}{2} Q(1, -\hat{t}) \quad (4)$$

as reference momenta, and Eq. (1) becomes

$$\tau_2^{ee} = \frac{1}{Q} \sum_k E_k \min\{1 - \cos \theta_k, 1 + \cos \theta_k\}, \quad (5)$$

where θ_k is the angle between \vec{p}_k and \hat{t} . The minimum divides all particles into the two hemispheres perpendicular to \hat{t} as shown in Fig. 1(a). For $\tau_2^{ee} \ll 1$, the total invariant mass in each hemisphere is much smaller than Q , so the final state contains two narrow jets. In this limit, $\tau_2^{ee} = 1 - T$, and a factorization theorem exists for $d\sigma/d\tau_2^{ee}$, which can be used to sum logarithms of τ_2^{ee} [4]. For a given jet algorithm with resolution parameter y , the value y_{23} marks the transition between 2 and 3 jets. Thus requiring $y_{23} \ll 1$ also vetoes events with > 2 jets.

$N = 0$ for Drell-Yan. Next, consider the isolated Drell-Yan process, $pp \rightarrow X \ell^+ \ell^-$ with no hard central jets, shown in Fig. 1(b). We now have ISR from the incoming partons, but no FSR from jets. From Eq. (3) we have

$$x_a E_{\text{cm}} = e^{+Y} \sqrt{q^2 + \vec{q}_T^2}, \quad x_b E_{\text{cm}} = e^{-Y} \sqrt{q^2 + \vec{q}_T^2}, \quad (6)$$

where q^2 and \vec{q}_T are the dilepton invariant mass and transverse momentum, and Y equals the dilepton rapidity. Now, $Q^2 = q^2 + \vec{q}_T^2$ and Eq. (1) becomes

$$\tau_0 = \frac{1}{Q} \sum_k |\vec{p}_{kT}| \min\{e^{Y - \eta_k}, e^{-Y + \eta_k}\}. \quad (7)$$

where $|\vec{p}_{kT}|$ and η_k are the transverse momentum and rapidity of p_k . The q_a and q_b dependence in Eq. (1) explicitly accounts for the boost of the partonic center-of-mass frame. For $Y = 0$, the minimum in Eq. (7) divides

all particles into two hemispheres perpendicular to the beam axis (analogous to \hat{t} above). For $Y \neq 0$, the hemispheres are boosted with their boundary now at Y , and the beam jet in the direction of the boost is narrower than the other, as depicted in Fig. 1(b). Contributions to τ_0 from large rapidities are exponentially suppressed by $|\vec{p}_{kT}| e^{-|\eta_k|} \approx 2E_k e^{-2|\eta_k|}$, so particles beyond the detector's rapidity reach give negligible contributions.

Beam thrust [5] is given by $\tau_B = \sqrt{1 + \vec{q}_T^2/q^2} \tau_0$. It is obtained by choosing $x_{a,b} E_{\text{cm}} = \sqrt{q^2} e^{\pm Y}$ in case q^2 and Y are measured rather than the longitudinal components $n_{a,b} \cdot q$ in Eq. (6). For $\tau_0 \ll 1$ the hadronic final state can only contain soft radiation plus energetic radiation in the forward directions, so $|\vec{q}_T| \ll Q$ and $\tau_B = \tau_0$. A factorization theorem for $d\sigma/d\tau_B$ at small τ_B was derived and used to sum logarithms of τ_B to NNLL in Refs. [5].

General case. For $pp \rightarrow XL(Nj)$ we have both ISR and FSR. We select candidate signal events by measuring L and running a jet algorithm to find the N signal jets and their momenta p_J . The conditions on the jets and L that define the signal are encoded in the cross section by a measurement function $F_N(\{p_J\}, L)$. Generically, F_N will enforce that there are at least N energetic jets that are sufficiently separated from each other and the beams. We now use the measured jet energies and directions to define the massless reference momenta q_J in Eq. (2),

$$E_J = p_J^0, \quad \hat{n}_J = \vec{p}_J/|\vec{p}_J|, \quad (8)$$

while q_a and q_b are given by Eqs. (2) and (3).

Taking the minimum in Eq. (1) combines the previous cases in Eqs. (5) and (7). It divides all particles into jet and beam regions that are unique for a given set of reference momenta and whose union covers all of phase space, as illustrated in Fig. 1(c). The boundary between any two neighboring regions is part of a cone and is such that the sum of the total invariant masses in the regions is minimized (or in case of a beam region the virtuality of the incoming colliding parton).

For events with small τ_N all jet algorithms should agree how energetic radiation is split up between the jets and beams, and only differ in their treatment of softer particles. Thus, they all give the same \hat{n}_J and E_J up to power corrections, while the split up of the soft radiation is determined by τ_N itself. Hence, the dependence of τ_N on the jet algorithm is formally power suppressed, $\tau_N^{\text{alg},1} = \tau_N^{\text{alg},2} + \mathcal{O}(\tau_N^2)$, as seen in Eq. (14) below.

To measure τ_N , we still rely on having a suitable jet algorithm to find the N signal jets but not more so than if we were not measuring τ_N . Imagine the jet size in the algorithm is chosen too small such that the algorithm divides what should be a single signal jet into several narrow jets [11]. In this case, the jet algorithm yields a poorly reconstructed signal irrespective of measuring τ_N .

Since the jet veto is now provided by τ_N , this situation can be avoided because we do not have to rely on the jet

algorithm to identify additional jets and so can use an algorithm that can be forced to always yield at most N jets. This is in fact the most natural thing to do when one is looking for N jets. Therefore, using τ_N as jet veto could also help improve the signal reconstruction.

Generalizations. We can generalize τ_N to

$$\tau_N^d = \sum_k \min\{d_a(p_k), d_b(p_k), d_1(p_k), \dots, d_N(p_k)\}, \quad (9)$$

where $d_m(p_k)$ can be any infrared-safe distance measure. In Eq. (1), $d_m(p_k) = 2q_m \cdot p_k/Q^2$ with

$$\begin{aligned} 2q_a \cdot p_k &= |\vec{p}_{kT}| Q e^{Y-\eta_k}, \\ 2q_J \cdot p_k &= |\vec{p}_{kT}| |\vec{q}_{JT}| (2 \cosh \Delta\eta_{Jk} - 2 \cos \Delta\phi_{Jk}). \end{aligned} \quad (10)$$

Here, $\Delta\eta_{Jk}$ and $\Delta\phi_{Jk}$ are the rapidity and azimuthal distances between q_J and p_k . If these are small, the factor in brackets reduces to the familiar $R^2 = (\Delta\eta)^2 + (\Delta\phi)^2$.

Different measures that are boost-invariant along the beam axis can be obtained by modifying the dependence on rapidity, $|\vec{q}_{JT}|$, and Q in Eq. (10). A geometric measure, which is independent of $|\vec{q}_{JT}|$, is

$$\begin{aligned} d_a(p_k) &= \frac{|\vec{p}_{kT}|}{Q} e^{Y-\eta_k}, & d_b(p_k) &= \frac{|\vec{p}_{kT}|}{Q} e^{-Y+\eta_k}, \\ d_J(p_k) &= \frac{|\vec{p}_{kT}|}{Q} (2 \cosh \Delta\eta_{Jk} - 2 \cos \Delta\phi_{Jk}). \end{aligned} \quad (11)$$

It evenly divides the area rather than invariant mass between neighboring regions, such that more energetic jets also get more invariant mass.

If L contains missing energy, then $x_{a,b}$ in Eq. (3) and thus Q and Y are not known. For Q , one can use any hard scale, like the $|\vec{q}_{JT}|$ of the hardest jet or leave it out, since it only serves as an overall normalization. In the beam measures $d_{a,b}(p_k)$ we can simply set $Y = 0$, which defines them in the hadronic center-of-mass frame.

N -jettiness does not split events into N , $N+1$, $N+2$, etc. jets like a traditional jet algorithm. But we can consider using τ_N to define an “exclusive N -jet algorithm” as follows: First, we use a geometric measure and find the directions \hat{n}_J and boost Y that minimize τ_N , analogous to finding \hat{t} for $e^+e^- \rightarrow \text{jets}$. This might actually allow one to get an estimate of Y even in the case of missing energy by exploiting the asymmetry in the beam jets. Second, we determine the jet energies by summing over the particles in each jet region. (To reduce the sensitivity to the underlying event and pile-up, one can weigh the sum over energies by the distance from \hat{n}_J .)

Factorization formula. We now use τ_N again as defined in Eq. (1). For $\tau_N \ll 1$, QCD ISR and FSR can be described in soft-collinear effective theory [6] at leading power by $N+2$ independent sectors for collinear particles close to each q_m with $m = \{a, b, J\}$ and a separate sector for soft particles. By power counting, J -collinear

particles are closest to q_J , so for the J -collinear sector

$$\sum_{k \in \text{coll}_J} \min_m \{2q_m \cdot p_k\} = \sum_{k \in \text{coll}_J} 2q_J \cdot p_k = s_J, \quad (12)$$

where (up to power corrections) s_J is the total invariant mass in the J -collinear sector. Similarly, the sum over the beam collinear sectors yields the total (transverse) virtuality of the colliding partons, t_a and t_b . Therefore,

$$\tau_N Q^2 = t_a + t_b + \sum_J s_J + \sum_{k \in \text{soft}} \min_m \{2q_m \cdot p_k\}. \quad (13)$$

The sum in the last term is now restricted to the soft sector. Combining Eq. (13) with the analyses in Refs. [5, 7] yields the factorization formula for N -jettiness [12]

$$\begin{aligned} \frac{d\sigma}{d\tau_N} = & \int dx_a dx_b \int d^4 q d\Phi_L(q) \int d\Phi_N(\{q_J\}) \\ & \times F_N(\{q_m\}, L) (2\pi)^4 \delta^4 \left(q_a + q_b - \sum_J q_J - q \right) \\ & \times \sum_{ij,\kappa} \text{tr} \hat{H}_{ij \rightarrow \kappa}(\{q_m\}, L, \mu) \prod_J \int ds_J J_{\kappa J}(s_J, \mu) \\ & \times \int dt_a B_i(t_a, x_a, \mu) \int dt_b B_j(t_b, x_b, \mu) \\ & \times \hat{S}_N^{ij \rightarrow \kappa} \left(\tau_N - \frac{t_a + t_b + \sum_J s_J}{Q^2}, \{q_m\}, \mu \right). \quad (14) \end{aligned}$$

Here, $\hat{H}_{ij \rightarrow \kappa}(\{q_m\}, L)$ contains the underlying hard interaction $i(q_a)j(q_b) \rightarrow L(q)\kappa_1(q_1) \cdots \kappa_N(q_N)$, where i, j , and κ_J denote parton types, and the sum over ij, κ is over all relevant partonic channels. It is a matrix in color space given by the IR-finite parts (in pure dim. reg.) of the squared partonic matrix elements in each channel. The N -body phase space for the massless momenta q_J is denoted $d\Phi_N(\{q_J\})$, and that for L by $d\Phi_L(q)$.

The inclusive jet and beam functions, $J_{\kappa J}(s_J)$ and $B_{i,j}(t_a, x_a, \mu)$, describe the final and initial state radiation emitted by the outgoing and incoming partons from the hard interaction. The latter also determine the momentum fractions $x_{a,b}$ of the colliding partons and are given by parton distribution functions $f_{i'}(\xi, \mu)$ as [5, 8]

$$B_i(t, x, \mu) = \sum_{i'} \int_x^1 \frac{d\xi}{\xi} \mathcal{I}_{ii'} \left(t, \frac{x}{\xi}, \mu \right) f_{i'}(\xi, \mu). \quad (15)$$

The $\mathcal{I}_{ii'}$ are perturbative coefficients describing collinear ISR, and at tree level $B_i(t, x, \mu) = \delta(t) f_i(x, \mu)$. The last term in Eq. (13) is the contribution to τ_N from soft particles in the underlying event. It is described by the soft function $\hat{S}_N^{ij \rightarrow \kappa}(\tau_N^{\text{soft}}, \{q_m\})$, which depends on the jet's angles $\hat{n}_l \cdot \hat{n}_m$ and energy fractions E_l/E_m . Like \hat{H} , it is a color matrix, and the trace in Eq. (14) is over $\text{tr}(\hat{H} \hat{S})$.

In Eq. (14), all functions are evaluated at the same renormalization scale μ . Large logarithms of τ_N in $d\sigma/d\tau_N$ are summed by first computing $\hat{H}(\mu_H)$, $J(\mu_J)$,

$B(\mu_B)$, $\hat{S}(\mu_S)$ at the scales $\mu_H \simeq Q$, $\mu_J \simeq \mu_B \simeq \sqrt{\tau_N} Q$, $\mu_S \simeq \tau_N Q$, where the functions contain no large logarithms, and then evolving them to the scale μ . This evolution is known analytically [9] and the required anomalous dimensions are already known to NNLL [5, 10], because we have inclusive jet and beam functions. NNLL also requires the $\mathcal{O}(\alpha_s)$ corrections for each function, which are known for J and B . The $\mathcal{O}(\alpha_s)$ hard function is determined by the one-loop QCD matrix elements. For $\tau_N \gg \Lambda_{\text{QCD}}/Q$, $\hat{S}(\mu_S)$ can be computed perturbatively and will be given in a future publication.

We thank C. Lee, K. Tackmann, and J. Thaler for comments and discussions. This work was supported by the Office of Nuclear Physics of the U.S. Department of Energy, under the grant DE-FG02-94ER40818.

- [1] CMS Collaboration, G. L. Bayatian *et al.*, J. Phys. G **34**, 995 (2007); ATLAS Collaboration, (1999), CERN-LHCC-99-15.
- [2] CDF and D0 Collaboration, T. Aaltonen *et al.*, Phys. Rev. Lett. **104**, 061802 (2010),
- [3] A. Banfi, G. P. Salam, and G. Zanderighi, JHEP **08**, 062 (2004); JHEP **06**, 038 (2010).
- [4] G. P. Korchemsky and G. Sterman, Nucl. Phys. B **555**, 335 (1999); S. Catani *et al.* Nucl. Phys. B **407**, 3 (1993); T. Becher and M. D. Schwartz, JHEP **07**, 034 (2008).
- [5] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, Phys. Rev. D **81**, 094035 (2010); arXiv:1002.2213.
- [6] C. W. Bauer, S. Fleming, and M. E. Luke, Phys. Rev. D **63**, 014006 (2000); C. W. Bauer *et al.*, Phys. Rev. D **63**, 114020 (2001); C. W. Bauer and I. W. Stewart, Phys. Lett. B **516**, 134 (2001); C. W. Bauer, D. Pirjol, and I. W. Stewart, Phys. Rev. D **65**, 054022 (2002); C. W. Bauer *et al.*, Phys. Rev. D **66**, 014017 (2002).
- [7] C. W. Bauer, A. Hornig, and F. J. Tackmann, Phys. Rev. D **79**, 114013 (2009).
- [8] S. Fleming, A. K. Leibovich, and T. Mehen, Phys. Rev. D **74**, 114004 (2006).
- [9] M. Neubert, Eur. Phys. J. C **40**, 165 (2005); S. Fleming *et al.*, Phys. Rev. D **77**, 114003 (2008); Z. Ligeti *et al.*, Phys. Rev. D **78**, 114014 (2008).
- [10] G. P. Korchemsky and A. V. Radyushkin, Nucl. Phys. B **283**, 342 (1987); S. Moch, J. A. M. Vermaseren, and A. Vogt, Nucl. Phys. B **688**, 101 (2004); G. Kramer and B. Lampe, Z. Phys. C **34**, 497 (1987), [Erratum-ibid. C **42**, 504 (1989)]; R. V. Harlander, Phys. Lett. B **492**, 74 (2000); S. Mert Aybat, L. J. Dixon, and G. Sterman, Phys. Rev. D **74**, 074004 (2006); A. Ferroglio *et al.*, JHEP **11**, 062 (2009).
- [11] This can be tested by comparing the total energy in each region defined by τ_N with the energy from the jet algorithm. If these are very different, but at the same time τ_N is small, then there are additional energetic particles near the signal jets that the algorithm should have included.
- [12] Here, F_N enforces distinct collinear sectors with $1 - \hat{n}_l \cdot \hat{n}_m \gg \tau_N$ and $E_m/Q \gg \tau_N$. We assume F_N only depends on the large components q_J of the jet momenta, $p_J = q_J[1 + \mathcal{O}(\tau_N)]$, and that L only couples to the QCD subprocess via a hard interaction. We also assume that Glauber gluons do not spoil this factorization.